Convergence proofs, convergence rates and stopping criterions for multi-modal or multi-objective evolutionary algorithms

Yann Bonnemay, Michèle Sebag, Olivier Teytaud

Convergence proofs for multi-modal multi-objective optimization

Y. Bonnemay*, M. Sebag*, O. Teytaud*,**
*TAO-inria, LRI, UMR 8623(CNRS - Universite Paris-Sud),
bat 490 Universite Paris-Sud 91405 Orsay Cedex France
**Artelys, www.artelys.com

May 13, 2005

Abstract

Evolutionary algorithms are population-based. It's an advantage for optimization problems in which a solution set, and not only a solution, is expected. Therefore, they are in particular relevant for

- multi-objective optimization (MOO), where the whole Pareto-front is interesting ;
- multi-modal optimization (MMO), where all local maxima are interesting ;
- multi-objective and multi-modal optimization (MOMMO).

As far as we now, state of the art convergence results are of two types :

- for stochastic algorithms, they concern convergence in distribution for multi-modal optimization (e.g. simulated annealing results),
- for deterministic algorithms, they concern the *inclusion* of accumulation points in the set of substationary points (which includes the Pareto front).

In this paper, we i) give convergence proofs for stochastic MMO, MOO and MOMMO algorithms (in the accumulation sense), ii) give convergence rates for stochastic MOO algorithms that can not be (without further hypothesis) extended to stochastic MMO or MOMMO algorithms, iii) provide convergence criterions for MOO problems. We then discuss the space complexity of population-based MOO, MMO and MOMMO algorithms.

We then prove upper and lower bounds on the space complexity of population based MOO and MOMMO algorithms.

1 Introduction

Multi-modal optimization is the research of the set of the global maxima of f or the set of the local maxima of f ([Li *et al.*, 2002],[Pena *et al.*, 2003]).

Multi-objective optimization ([Bhattacharyya, 2000, Miettinen, 1999, Deb, 2001]) is the research of the set $\{x; \nexists y; (\forall i\ f_i(x) \leq f_i(y)) \wedge (\exists i\ f_i(x) < f_i(y))\}$ which is called the Pareto front. It is of increasing importance in data mining ([Francisci *et al.*, 2003, Ghosh and Nath, 2004, de la Iglesia *et al.*, 2003, Laumanns *et al.*, 2002]). The diversity of the population is a main goal of MOO ([Toffolo *et al.*, 2005]).

Multi-objective multi-modal optimization ([Sebag *et al.*, 2005]) is the research of the set of x such that locally x is a multi-objective optimum. It is in particular relevant for spatial and temporal data-mining ([Roddick and Spiliopoulou, 2002, Shekhar *et al.*, 2003]). MOMMO problems have been formulated in [Sebag *et al.*, 2005] as dominance-based optimization problems, for which evolutionary algorithms are relevant.

In the sequel, we :

- define a non-standard relation of dominance, that we call "β-dominance", for the equivalent of dominance naturally associated to both MMO and MOMMO frameworks ;
- provide a convergence proof for a class of relations called α-dominance including Pareto-dominance ;
- adapt this proof to the class of relations of β-dominance ; this class includes MMO and MOMMO, both in the sense of the research of the set of global maxima and in the sense of the research of the set of local maxima ;
- prove convergence rates for the Pareto-dominance ;
- prove upper and lower bounds for the size of the population.

The convergence proofs of items 2 and 3 only assume properties of the dominance relations.

Section 2 summaries the state of the art. Section 3 provides the notations. Section 4 defines an evolutionary algorithm for MMO, MOO or MOMMO problems and provides the hypothesis. Section 5 provides the results (5.1 : MOO ; 5.2 : MOMMO or MMO ; 5.3 : MOO with convergence rates). Section 6 concludes.

2 State of the art

Many papers have been devoted to multi-objective or multi-modal optimization, some of them with deterministic methods (see [Miettinen, 1999]), and some others with evolutionary algorithms ([Deb, 2001]).

Convergence proofs for iterative deterministic algorithms (see e.g. [Miettinen, 1999, chap.3]) for multi-objective optimization are restricted to *inclusions* of the form "the accumulation points are substationary" or results like this. The reverses inclusion does not hold for these algorithms. An advantage is the fast (superlinear) convergence rate.

Convergence proofs for stochastic algorithms in the case of multi-modal optimization mainly concern the simulated annealing ([Hajek, 1988]). These results state convergence in distribution to the set of global maxima. As far as we know, no convergence results have been stated for the convergence to the set of local maxima. [Garnier *et al.*, 2002] provides estimates of the number of local maxima in combinatorial optimization.

3 Notations

We consider P a distribution of probability on $[0,1]^D$. We note $B(c,r)$ the ball of center c and radius r. If A has cardinal $|A| \leq k$, we note A^k a set of cardinal k, containing A. A^k is not precisely defined : the important point for us is that i) it has cardinal k ii) it contains A iii) at least one of the $k-|A|$ other points is randomly drawn, in an independent manner, according to law P. We note $d(.,.)$ the euclidean distance between elements or subsets of $[0,1]^D$ (the distance $d(A,B)$ between subsets A and B is $\inf_{a\in A, b\in B} d(a,b)$).

We now consider relations between elements of $[0,1]^D$. We note "a α-dominates b" a partial order on elements in $[0,1]^D$. "a strictly α-dominates b" notes "a α-dominates b and $a \neq b$". We note "a β-dominates b" a relation. "a strictly β-dominates b" if "a β-dominates b and $a \neq b$". We note strict α-dominance by $a \succ b$ and strict β-dominance by $a \succ^\beta b$. We note $a \succ^\beta_k b$ the fact that there exists $x_1, \ldots, x_k$ such that $a \succ^\beta x_1 \succ^\beta x_2 \succ^\beta x_3 \ldots \succ^\beta x_k = b$. We note $A \succ B$ if $\forall (x,y) \in A \times B, x \succ y$. For the consistency of notation, we note $a \succ^\beta_k b$ the fact that $a = b$.

A typical and important case of relation of α-dominance is the Pareto-dominance : $a \succ b$ if and only if $\forall i, a_i \geq b_i$ and $\exists i, a_i > b_i$.

We define $PF(A)$ the set of elements in A that are not strictly dominated by any element in A. We define $PF^\beta(A)$ the set of elements in A that are not strictly β-dominated by any element in A. We note $PF = PF(support(P))$ and $PF^\beta = PF^\beta(support(P))$ where $support(P)$ is the support of P, i.e. the smallest (for the inclusion) closed set whose complementary has probability 0.

4 Evolutionary algorithm and hypothesis

We consider the following evolutionary algorithm :

- P_0 is the empty set.
- $P_n = (PF(P_{n-1}))^{N_n}$ for $n \geq 1$ (case of α-dominance),
- $P_n = (PF^\beta(P_{n-1}))^{N_n}$ for $n \geq 1$ (case of β-dominance),

We can identify P_n with a vector of $D \times N_n$ numbers in $[0,1]$ (so $P_n \in [0,1]^{DN_n}$).

Define $PF_n = PF(P_n)$ and $PF^\beta_n = PF^\beta(P_n)$. In the sequel, the accumulation denotes the accumulation of the sequence enumerating all points in PF_0, PF_1, PF_2, ... (resp. PF^β_i)

Axiomatic definitions of the α-dominance and the β-dominance

We assume that $\succ$ is a partial order and that :
H0 : N_n is at least the cardinal of $PF(P_n)$ plus 1 (resp. $PF^\beta(P_n)$ plus one) ;
H1 : A^k for k larger than the cardinal of A contains at least one point randomly drawn, in an independent manner, according to law P ;
H2 : with x a random variable with probability law P, $\forall p \in PF$ (resp. PF^β) $\forall \epsilon > 0$ $P(d(x,p) < \epsilon) > 0$;
H3 : $\forall y \notin PF$ (resp. PW^β), $\exists x \in PF$ (resp. PF^β), $\exists \epsilon > 0, B(x,\epsilon) \succ B(y,\epsilon)$.
H4 : $x \in PF$ implies that for any ϵ, for some $\epsilon' > 0$, $d(x,y) < \epsilon' \Rightarrow$ almost surely $d(a,x) < \epsilon \vee \neg(a \succ y)$ where a is a random variable with probability law P.
H4$^\beta$: $x \in PF^\beta$ implies that for any $\epsilon_0 > 0$, for some $0 < \epsilon < \epsilon_0$, $(d(x,z) \leq \epsilon \wedge a \succ_k^\beta z) \Rightarrow (d(x,a) \leq \epsilon_0)$.
Condition H0 states that at least one point is added at each step to the current population $PF(P_n)$ (this point might be discarded during the selection) ; condition H1 states that this point is randomly drawn. H3 states that the domination is "regular" (in particular, it forbids substationary points out of the Pareto front). H4$^\beta$ states a condition analoguous to the local convexity hypothesis in optimization ; it states the existence of basins of attraction.

Let's now verify that the two other axioms are reasonnable, and in particular that for the Pareto-dominance they are verified :

Lemma :

- H2 is a consequence of the definition of PF (resp. PF^β).
- H4 holds in the classical Pareto-dominance.

Proof :

- let's now prove the case of H2. Consider $p \in S = \mathit{support}(P)$. Assume, in order to get a contradiction, that H2 does not hold. Then, $\exists p \in PF, \exists \epsilon > 0; P(B(p,\epsilon)) = 0$. Then, $P(S \setminus B(p,\epsilon)) = 1$, which is a contradiction with the definition of S.
- we now prove the case of H4. Assume that H4 does not hold. Then, for some $\epsilon > 0$, for any $\epsilon' > 0$, there exists $a_{\epsilon'}$ in $\mathit{support}(P)$ such that $a_{\epsilon'} \succ y$, where $y \succ x + \epsilon'$ (where $x + \epsilon'$ is the coordinate-per-coordinate sum), and therefore $a_{\epsilon'} \succ x + \epsilon'$. We then consider an accumulation point a_0 of $a_{\epsilon'}$ as $\epsilon' \to 0$. $a_0 \in \mathit{support}(P)$ (as $\mathit{support}(P)$ is closed) and in the classical Pareto-framework $a_0 \succ x$ (by extension to the limit, $\forall i \in [[1,D]]$, $\pi_i(a_0) \geq \pi_i(x)$ for the projection π_i on the i^{th} coordinate, and $d(a_0,x) \geq \epsilon$). So, $a_0 \succ x$ and $a_0 \in \mathit{support}(P)$, which is a contradiction with $x \in PF$.

$\blacksquare$

We do not consider any constraint. They can be naturally added by ∞-penalization of the fitness when a constraint is violated.

5 Proofs

We here provide the theorems and proofs. Section 5.1 is the MOO case and is only based upon properties of the α-dominance relation. Section 5.2 is the MMO and the MOMMO case ; it is only based upon properties of the β-dominance relation. Section 5.3 provides convergence rates ; it uses less general definitions and is only relevant for Pareto-dominance.

5.1 α-dominance (MOO problems)

Theorem 1 : (case of α-dominance) Assume H0, H1, H2, H3, H4 and assume that $\succ$ is a partial order. Then the accumulation is equal to PF.

Proof :

Note *acc* the accumulation.

Step 1: after H2, H4 and by transitivity of the α-dominance, almost surely $x \in PF \Rightarrow \forall \epsilon \exists n_0 \forall n \geq n_0 PF_n \cap B(x, \epsilon) \neq \emptyset$.

Step 1 derives from the fact that for any $x \in PF$, for any $\epsilon > 0$,

1. by H4, for some ϵ', all points y at distance $\leq \epsilon'$ of x verify that almost surely $\neg(a \succ y)$ or $d(a, x) \leq \epsilon$ (where a is a random variable with law P).
2. H2 implies that x is approximated by one random point y generated at some epoch n_0, within distance ϵ'.
3. by the selection rule, all future generations will contain y or a point dominating y, and necessarily by item (1) this point is at distance at most ϵ of x.

Step 2: By transitivity and by the selection rule, $a \in PF(P_n) \Rightarrow \forall k > 0, \forall x \in PF(P_{n+k}), a \not\succ x$.

Step 3: After H2 and H3, $y \notin PF \Rightarrow \exists n; \exists a \in PF(P_n); \exists \epsilon; a \succ B(y, \epsilon)$. Proof : consider $y \notin PF$.

- by H3, $\exists x \in PF, \exists \epsilon > 0, B(x, 2\epsilon) \succ B(y, 2\epsilon)$.
- by H2, $\exists n, \exists a \in PF(P_n), d(a, x) \leq \epsilon$; a and ϵ have the expected properties.

Step 4: After steps 2 and 3, $y \notin PF \Rightarrow \forall k, PF(P_{n+k}) \cap B(y, \epsilon) = \emptyset \Rightarrow y \notin acc$.

Step 5: After step 4, $acc \subset PF$.

Step 6: Assume (to get a contradiction) that $x \in PF$ and $B(x, \epsilon) \cap acc$ is empty. Then $x \in PF$ and for some n_0, $n \geq n_0 \Rightarrow B(x, \epsilon) \cap PF_n = \emptyset$. This is a contradiction with step 1 ; therefore, for any $x \in PF$, for any $\epsilon > 0$, $B(x, \epsilon) \cap acc$ is non-empty.

Step 7: Step 6 implies that $d(x, acc) = 0$. *acc* is closed ; therefore, after step 6, any $x \in PF$ is in *acc* and so $PF \subset acc$.

Step 8: After steps 7 and 5, $acc = PF$.

■

Note the following

Corollary 2 : The theorem above states that $PF = \{x; \inf_{m\geq n} d(x, PF_m) \to 0\}$ (limit for $n \to \infty$). Thanks to step 1, indeed, the following stronger property holds :

$$PF = \{x; d(x, PF_n) \to 0\}$$

Corollary 3 : As PF is closed (as it is equal to *acc*) and bounded,

$$\sup_{x\in PF} d(x, PF_n) \to 0$$

Interpretation : The population converges, in all relevant senses, to the expected set, under mild assumptions on the α-dominance-relation (without convergence rates). The result includes MOO problems, and is only based on properties of $\succ$ (it is true for all relations $\succ$ satisfying the axioms in section 4).

5.2 β-dominance (MMO and MOMMO)

Theorem 4 : (case of β-dominance)

Assume that $\succ^\beta$ is a relation, and assume H0, H1, H2, H3 and H4$^\beta$. Then the accumulation is equal to PF^β.

Proof :

We cannot directly use the proof of theorem 1. The first trouble is that in the case of β-dominance the proof of step 1 does not hold here. On the other hand the following modified proof holds :

Step 1 : $x \in PF^\beta \Rightarrow \forall\epsilon_0 > 0, \exists n_0 \forall n \geq n_0; PF_n^\beta \cap B(x, \epsilon_0) \neq \emptyset$.

This is a consequence of H4$^\beta$, that is shown as follows :

1. consider some $x \in PF^\beta$ and $\epsilon_0 > 0$;
2. consider ϵ provided by H4$^\beta$;
3. by H2, almost surely a point y within distance ϵ of x is generated at some epoch n_0.
4. by construction, for any epoch $n \geq n_0$, either y, or a point z such that $z \succ_k^\beta y$, belongs to PF_n^β.
5. by H4$^\beta$, z verifies the property $x \succ^\beta z$ and $d(x, z) \leq \epsilon$.

Therefore, for any $n \geq n_0$, $PF_n^\beta \cap B(x, \epsilon) \neq \emptyset$.

Steps 6-7 hold as in theorem 1, and lead to the claim $PF^\beta \subset acc$. Steps 2-4 are replaced by the following steps proving $y \notin PF^\beta \Rightarrow y \notin acc$:

1. Assume that $y \notin PF^\beta$. Therefore, by H3, y is dominated by some $x \in PF^\beta$;
2. use H3 to choose ϵ_0 so that $B(x, \epsilon_0) \succ^\beta B(y, \epsilon_0)$;
3. consider $\epsilon > 0$ provided by H4$^\beta$.
4. by H2, x is approximated within distance $\leq \epsilon$ by some a in epoch n_0 ($a \in P_{n_0}$, $d(a, x) \leq \epsilon$ and $d(a, x) < \epsilon$);

5. for $n \geq n_0$, PF_n^β contains one point b such that for some k, $b \succ_k^\beta a$;
6. by H4$^\beta$ this b lies within distance ϵ_0 to x ($d(x,b) < \epsilon_0$) ;
7. by item 3 above, b dominates $B(y, \epsilon_0)$ and thus for $n \geq n_0$ the n^{th} generation can not contain any element in $B(y, \epsilon_0)$.

Step 8 can be used as in the proof of theorem 1. ■

Therefore, corollaries 2 and 3 are available as well for β-dominance :

Corollary 5: Thanks to step 1,

$$PF^\beta = \{x; d(x, PF_n) \to 0\}$$

and by compactness,

$$\sup_{x \in PF^\beta} d(x, PF_n^\beta) \to 0$$

Interpretation : The population converges, in all relevant senses, to the expected set, under mild assumptions on the β-dominance-relation. The result is only based on properties of $\succ^\beta$ (it is true for the class of all relations satisfying the axioms in section 4) and includes MMO and MOMMO problems.

5.3 Convergence rates : the classical Pareto-framework

Let's assume H5 : $a \succ b$ if and only if $\forall i,\ a_i \geq b_i$, and $\exists i,\ a_i > b_i$ (classical Pareto-dominance in the space of criterions), and H6 : P has bounded density wrt the Lebesgue measure.

Then, define m (possibly random variable implicitly depending upon n) the number of random drawns in $P_1, \ldots, P_n$. m is typically deterministic and equal to n if $\forall n, N_n = |PF(P_{n-1})| + 1$.

Note $X = (x_1, \ldots x_m)$ the set of the m first random drawns (all of them are independent with law P) and P_X its law. Consider x a random variable with probability law P. Note R_z the rectangle $[z_1, \infty] \times [z_2, \infty] \times \cdots \times [z_d, \infty]$.

We note as a first measure of quality of PF_n the (random) quantity $L_m = \sup_{y \in PF_n} P(x \succ y)$ where x is a random variable with probability law P. This is a form of Pareto-rank.

We now prove bounds on L_m, as follows.

Step 1 : Thanks to classical results of VC-theory (see e.g. [Devroye et al, 1997]), the VC-dimension of $R = \{R_y; y \in [0,1]^D\}$ is D.

Step 2 : we can apply VC-dimension bounds on uniform deviations to prove that rectangles as above that contain none of the m points $x_1, \ldots, x_m$ have a low measure for P. We will use a standard VC-inequality [Devroye et al, 1997, Th. 12.7], which implies that for any $y \in [0,1]^D$, for any $\epsilon \geq 0$,

$$P_X(\forall y \text{ s.t. } |X \cap R_y| = 0 \Rightarrow \underbrace{(P(x \succ y))}_{=L_m} > \epsilon) \leq 2S(2m)2^{-m\epsilon/2}$$

where $S(2m)$, $2m^{th}$ shattering coefficient of R, is upper-bounded by $\sum_{i=0}^{D} \frac{(2m)!}{(2m-i)!i!} \leq (2em/D)^D$ if $m \geq D$.

This leads to the following

Theorem 5 : (convergence rate of multi-objective optimization) If $m \geq D$, then

$$P_X(L_m \geq \epsilon) \leq 2(2em/D)^D 2^{-m\epsilon/2}$$

what leads to a weak $O(D\log(m)/m)$ convergence ($O(.)$ can be developped in explicit constants by usual algebra).

Interpretation : The result is independent of the distribution. This is not a fully satisfactory result, as it only ensures the quality of elements in the front, instead of the quality of the whole front. This leads to another criterion $L'_m = P(\neg\exists y \in PF_n; y \succ x)$.

Theorem 6 : (global convergence rate of multi-objective optimization) Consider a sequence $M(m) = o(m)$ of integers. If $m \geq M(m)D$, then

$$P_X(|PF_n| \leq M(m) \wedge L'_m \geq \epsilon) \leq 2(2em/(DM(m)))^{DM(m)} 2^{-m\epsilon/2}$$

i.e. within risk 5%, if $|PF_n|$ has size $\leq M(m)$, $L'_m \leq \epsilon$ with

$$2(2em/(DM(m)))^{DM(m)} 2^{-m\epsilon/2} = 0.05$$

what is a weak $O(\log(m)DM(m)/m)$ convergence of L'_m to 0, provided that $|PF_n| \leq M(m)$ ($O(.)$ can be developped in explicit constants by usual algebra).

Interpretation : Design a relevant sequence $M(m)$ (we will see below how to choose this sequence). Then, perform the search, including m random generations with law P, and generating an approximation of the Pareto-Front dominating all the random generations. If the size of the population defining the approximation of the Pareto-Front is bounded above by $M(m)$, then L'_m is bounded above as explained. This is in the spirit of sparsity bounds in learning. The result is independent of the distribution.

Second interpretation : The Pareto-Front is significant whenever its size is small in front of the number of random search performed.

Third interpretation : Assume that for the sequence $M(m)$, $|PF_n|$ is very likely to be lower than $M(m)$, and $M(m) = o(m)$. Then, the theorem provides a convergence rate (neglecting logarithmic factors weakly $L'_m = O(Dm^{-1/D})$ with $M(m) = O(m^{(D-1)/D})$ as explained below).

Remark (under some hypothesis on the distribution and on the algorithm) : This bound has no interest if $|PF_n| > M(m)$ or $M(m) = \Omega(m)$. So, we must ensure that for n sufficiently large, for some $M(m) = o(m)$, $|PF_n| \leq M(m)$ with large probability. We will indeed show that if $P_n = PF(P_{n-1})^{|N_n|}$ with $N_n = |PF(P_{n-1})| + 1$ (i.e., just add one random point and update the Pareto-front), then a small $M(m)$ is an upper bound on $|PF_n|$ if n is sufficently large. This is indeed a consequence of theorem 5 :

- Theorem 5 states that with high probability, any y in PF_n verifies $P(x \succ y) = O(D\log(m)/m)$ where x is a random variable with law P.
- Assume that $P(x \succ y) = \Omega(d(y, PF)^D)$ (smoothness of the distribution). Then, $d(y, PF) = O(\sqrt[D]{D\log(m)/m})$.

- Therefore, with high probability, $|PF_n| \leq \sum_{i=1}^{m} \chi_{d(x_i,PF)\leq K\sqrt[D]{D\log(m)/m}}$ for some K, where χ_E the characteristic function of event E.
- Assume $P(d(x_1,PF) \leq \epsilon) = \theta(\epsilon)$. This hypothesis is not fully natural, but other assumptions like $\theta(\epsilon^r)$ lead to very related results. This implies that with high probability, $|PF_n| \leq B(m, K_2\sqrt[D]{D\log(m)/m})$ where $B(m,p)$ is the sum of m independent variables equal to 1 with probability p and 0 otherwise.
- $B(m, K_2\sqrt[D]{D\log(m)/m})$ is bounded above by $K_3 m^{(D-1)/D}\sqrt[D]{D\log(m)}$ with high probability (ie, for any $\delta > 0$, there exists K_3 such that $P(B(m, K_2\sqrt[D]{\log(m)/m}) > K_3 m^{(D-1)/D}\sqrt[D]{D\log(m)}) \leq \delta$.
- Hence, $M(m) = K_3 m^{(D-1)/D}\sqrt[D]{D\log(m)}$ satisfies the requirement that $P_X(|PF_n| > M(m))$ is small (smaller than a probability which only depends upon K_3 and goes to 0 as K_3 increases).

After this remark showing that $M(m)$ can be chosen of the order of $K_3 m^{(D-1)/D}\sqrt[D]{D\log(m)}$, and thus that L'_m converges roughly as $1/\sqrt[D]{m}$, we can now turn our attention to the proof of theorem 6.

Proof :

Step 1 : The set $\{\{x; \{y_1,\ldots,y_{M(m)}\} \succ x\}; (y_1,\ldots,y_{M(m)}) \in ([0,1]^D)^{M(m)}\}$ has VC-dimension upper-bounded by $DM(m)$ (see e.g. [Devroye et al, 1997, chap. 13]).

Step 2 : Therefore, by [Devroye et al, 1997, Th 12.7] again, sets of this form containing the m random points $x_1, \ldots, x_m$ contain, with high probability, a measure at least $1-\epsilon$ for P :

$$P_X(L'_m \geq \epsilon) \leq 2(2em/M(m)D)^{DM(m)} 2^{-m\epsilon/2}$$

■

Interpretation : This is quite slow, in front of superlinear deterministic methods. But we here ensure convergence to the *whole* Pareto-Front.

Unfortunately, this has no straightforward adaptation to the case of β-domination. The main tool of the proof is the fact that when a generation contains a point, all future generations dominate all the points dominated by this point. This is due to transitivity, which does not hold for β-domination. This is not very surprising as the local nature of multi-modal optimization naturally leads to slow convergence results.

We now turn our attention to the space complexity of population-based optimization, an issue that becomes important in evolutionary computation ([Gao *et al.*, 2005]) and in particular in set-oriented evolutionary computation. Define δ the dimension (if it exists) of the set $E \subset [0,1]^D$ as follows :

$$\exists K_1, K_2; \forall 1 > \epsilon > 0;\ \text{the smallest } F \subset E \text{ such that } E \subset \bigcup_{x\in F} B(x,\epsilon)$$

$$\text{verifies } |F| \in [K_1(\frac{1}{\epsilon})^\delta, K_2(\frac{1}{\epsilon})^\delta]$$

with δ minimal in that case. δ is the natural dimension of E for many definitions and for smooth classes of sets. In particular, in MOO, it is (under mild smoothness hypothesis) the number of criterions minus 1.

Then, assume that P covers E within precision ϵ in the following sense

$$E \subset \bigcup_{x \in P} B(x, \epsilon)$$

Note that this is a non-restrictive definition as it is verified if $\bigcup_{x \in P} B(x, \epsilon) = [0,1]^D$. Results proved below on the basis of this lemma are therefore true even for approximations that are much "bigger" than the target set. Anyway, it is sufficient for deriving the following property :

Lemma 7 : If P approximates a set of dimension δ within precision ϵ, then

$$|P| \geq K_1 (\frac{1}{\epsilon})^\delta$$

Interpretation : This simple results stands alone : it shows that for a good approximation of a set, we need a population of size polynomial in ϵ but exponential in the dimension of the set.

We now consider the rate of convergence of the algorithm. This shows that when no-compression of the target-set E is possible (such compression occurs in particular in linear case, see e.g. MOLP problems, [Evans et al, 1973]), the space complexity of the algorithm is necessarily huge, and it increases exponentially in δ.

We now consider the application of lemma 7 to the case of L'_m decreasing as explained in theorem 6. In our convergence rate above, in the case detailed in the remark after theorem 6 :

- L'_m decreases (in the probabilistic sense - it's a weak convergence) as $DM(m)\log(m)/m$.
- This implies, if the density and the Pareto front are smooth enough, that it is approximated within precision $O((DM(m)\log(m)/m)^{1/D})$ (otherwise else, an area[1] larger than $\Omega(DM(m)\log(m)/m)$ is not covered by PF_n),
- and then by lemma 7 the complexity is $\Omega((m/(D\log(m)M(m)))^{\delta/D})$.

This implies that the population size is at least of the order, in the case detailed in the remark after theorem 6, $m^{(D-1)/D^2}$ (neglecting logarithmic factors).

Thus we have shown the following

Theorem 8 : Consider the MOO case with Pareto-Front of dimension $D-1$ and with assumptions as in the remark after theorem 6. The random search $PF_n = PF(x_1, \ldots, x_m)$ has size weakly[2] $|PF_n| = \Omega(m^{(D-1)/D^2})$ (neglecting logarithmic factors).

Interpretation : The MOO (and MOMOO as well) problems converging as quickly as the random search in non-trivial cases have large space-complexity of the order $m^{(D-1)/D^2}$.

[1] area for P.

[2] Weak $a = \Omega(b)$ means that $\lim_n P(a \leq Kb)$ exists for any K.

6 Conclusion

We have provided an axiomatic definition of two dominance relations and have proved :

- that the axiomatic definitions of α-dominance and β-dominance are naturally verified under mild hypothesis on usual MMO, MOO or MOMMO problems ;
- population-based convergences for MMO, MOO and MOMMO problems which generalize the state of the art (in particular, our general framework based upon axiomatic properties of dominations includes research of local maxima as in the research of spatio-temporal patterns, and not only global maxima) ;
- convergence rates in a MOO framework, with convergence indicators based on the population size ;
- upper and lower bounds on the space complexity if the dimension of the target set is > 0 (what is the case in non-trivial MOO problems) ;
- we discussed the fact that the convergence rate could not be (at least without further hypothesis) extended to the MOMMO or MMO frameworks.

The main limits of our work is that the mathematical treatement includes many evolutionary algorithms, but this diversity is due to the fact that we mainly use two features, namely i) the population-based nature of the algorithm ; ii) the random diversification. We unfortunately do not take into account fine genetic operators or heuristics.

References

[Back, 1995] T. Bäck. *Evolutionary Algorithms in theory and practice.* New-York:Oxford University Press,1995.

[Bhattacharyya, 2000] S. Bhattacharyya, Evolutionary algorithms in data mining: multi-objective performance modeling for direct marketing. Proceedings of KDD, pp 465-473, 2000.

[Deb, 2001] K. Deb. *Multi-Objective Optimization Using Evolutionary Algorithms.* John Wiley, 2001.

[Devroye et al, 1997] L. Devroye, L. Györfi, and G. Lugosi, A probabilistic Theory of Pattern Recognition, Springer, 1997.

[Evans et al, 1973] J.P. Evans, and R.E. Steuer : A Revised Simplex Method for Linear Multiple Objective Programs, Mathematical Programming 5, 54-72, 1973.

[Francisci *et al.*, 2003] D. Francisci, and M. Collard, Multi-Criteria Evaluation of Interesting Dependencies according to a Data Mining Approach, Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003), Vol. 3, pp. 1568-1574, IEEE Press, Canberra, Australia, 2003.

[Gao *et al.*, 2005] Y. Gao and J.C. Culberson, Space Complexity of Estimation of Distribution Algorithms Vol. 13, Issue 1. 2005.

[Garnier *et al.*, 2002] J. Garnier, and L. Kallel, Efficiency of local search with multiple local optima, SIAM J. Discrete Math., Vol. 15, pp. 122–141, 2002.

[Ghosh and Nath, 2004] A. Ghosh and B. Nath. Multi-objective rule mining using genetic algorithms. *Inf. Sci.*, 163(1-3):123–133, 2004.

[Goldberg, 2002] D. Goldberg. *The Design of Innovation: Lessons from and for Genetic and Evolutionary Algorithms.* MIT Press, 2002.

[Hajek, 1988] B. Hajek. Cooling Schedule for Optimal annealing. Mathematics of Operations Research 13, 311-329, 1988.

[de la Iglesia *et al.*, 2003] B. de la Iglesia, M.S. Philpott, A.J. Bagnall, and V.J. Rayward-Smith, Data Mining Using Multi-Objective Evolutionary Algorithms, Proceedings of the 2003 Congress on Evolutionary Computation (CEC'2003), Vol. 3, pp. 1552-1559, IEEE Press, Canberra, Australia, 2003.

[Keim *et al.*, 2004] D. A. Keim, J. Schneidewind, and M. Sips. Circleview: a new approach for visualizing time-related multidimensional data sets. In *Proc. of Advanced Visual Interfaces*, pages 179–182. ACM Press, 2004.

[Laumanns *et al.*, 2002] M. Laumanns, L. Thiele, K. Deb, and E. Zitsler. Combining convergence and diversity in evolutionary multi-objective optimization. *Evolutionary Computation*, 10(3):263–282, 2002.

[Li *et al.*, 2002] J.-P. Li, M. E. Balazs, G. T. Parks, and P. J. Clarkson. A species conserving genetic algorithm for multimodal function optimization. *Evolutionary Computation*, 10(3):207–234, 2002.

[Miettinen, 1999] K.M. Miettinen, Nonlinear Multiobjective Optimization, Kluwer Academic Publishers, Boston, Massachusetts, 1999.

[Roddick and Spiliopoulou, 2002] J.F. Roddick and M. Spiliopoulou. A survey of temporal knowledge discovery paradigms and methods. *IEEE Trans. on Knowledge and Data Engineering*, 14(4):750–767, 2002.

[Sebag *et al.*, 2005] M. Sebag, N. Tarrisson, O. Teytaud, J. Lefevre, S. Baillet. Multi-Objective Multi-Modal Optimization Approach
for Mining Stable Spatio-Temporal Patterns. , IJCAI 2005.

[Pena *et al.*, 2003] J.-M. Pena, J.A. Lozano, P. Larranaga, Globally Multimodal Problem Optimization Via an Estimation of Distribution Algorithm Based on Unsupervised Learning of Bayesian Networks, *Evolutionary Computation*, 13(1):43-66, 2003.

[Shekhar *et al.*, 2003] S. Shekhar, P. Zhang, Y. Huang, and R. R. Vatsavai. Spatial data mining. In H. Kargupta and A. Joshi, eds, *Data Mining: Next Generation Challenges and Future Directions*. AAAI/MIT Press, 2003.

[Toffolo *et al.*, 2005] A. Toffolo, E. Benini, Genetic Diversity as an Objective in Multi-Objective Evolutionary Algorithms, *Evolutionary Computation*, 11(2):151–168, 2003.